

Costa, M. A., Luegi, P. (2009). Complexidade linguística e processamento referencial. In Dermeval da Hora (Org.), *Anais do VI Congresso Internacional da ABRALIN*. CD-ROM, pp. 2127-2135.

COMPLEXIDADE LINGÜÍSTICA E PROCESSAMENTO REFERENCIAL<sup>1</sup>

Maria Armanda Costa – FLUL  
Paula Luegi – CLUL

### Introdução

Um dos problemas mais interessantes a enfrentar na investigação sobre o processamento da linguagem verbal, particularmente no processamento de frases, é a possibilidade de identificar indicadores de complexidade linguística que se possam relacionar com custos de processamento e, portanto, com variações na compreensão.

A complexidade pode associar-se a factores que relevam do léxico, que podem ter a ver com a própria estrutura fonológica da palavra e a sua representação ortográfica com impacto no reconhecimento visual, ou com questões de frequência lexical, ou aspectos de vizinhança entre palavras que aceleram ou retardam o acesso ao significado, ou mesmo com graus de especialização do vocabulário. Pode também a complexidade relevar de aspectos do discurso, com formas mais ou menos canónicas de apresentar a informação – tipos de textos mais acessíveis do que outros – ou dos recursos retóricos usados, ou das próprias convenções de escrita, como a pontuação ou o *layout*.

A complexidade que por agora mais nos interessa é aquela que se prende com as estruturas gramaticais, que varia de acordo com as diferentes propriedades sintácticas das estruturas. Assim, tomando por referência estruturas básicas não marcadas, poder-se-ão tomar como indicadores de complexidade quaisquer propriedades sintácticas que resultem da alteração dos padrões básicos de ordem de palavras, de processos de ligação de frases, nomeadamente de níveis de encaixe em construções de subordinação, de construções de elipse, ou da existência de dependências a longa distância. Vejam-se os resultados encontrados no processamento de frases marcadas informacionalmente que apresentam movimento de constituintes em línguas românicas, como o Português, o Italiano e o Francês (Costa 2005, 2006; De Vincenzi 1991; Kail 1989), frases com dependências a longa distância, como nas construções WH-, interrogativas e relativas (Vasconcelos 1996; Gibson 1998; Warren e Gibson 2002), em construções de subordinação em função dos níveis de integração dos vários tipos de subordinadas (Gayraud e Martinie 2008). É também factor potenciador de complexidade a existência de ambiguidade estrutural que coloca o *parser* em situação de decisão sobre estruturas competitivas. Em Português Europeu, veja-se o trabalho de Costa (2005), sobre frases com verbos transitivos e inversão do sujeito, ou, sobre o Português Europeu e o Português Brasileiro, os trabalhos sobre o processamento de NPs complexos com relativas ambíguas, em Maia *et al.* (2007), por exemplo.

Um caso interessante para o estudo do processamento *on-line* de frases sob condições de sobrecarga cognitiva causada por complexidade linguística é o da necessária atribuição de conteúdo a expressões referencialmente dependentes. As condições de recuperação de um antecedente para uma dada expressão anafórica são afectadas por uma variedade de factores: condições de ambiguidade em que várias entidades se colocam como possíveis identificadores de um mesmo pronome (Costa, Faria e Matos 1997; Faria e Kail 2004; Costa 2005), a natureza da expressão anafórica que pode indiciar o maior ou menor grau de activação da entidade referida (Ariel 1996; Almor 2000; Leitão 2005), a variação da distância entre antecedente e anáfora (sentido lato), que pode afectar capacidades de armazenamento da memória de trabalho, assim como as propriedades sintácticas do material inserto entre antecedente e pronome.

Neste estudo, focamo-nos no processamento referencial que pode complexificar-se por uma conjugação de factores sintácticos e semânticos que podem pesar nos recursos cognitivos movidos para a atribuição da referência no processamento *on-line*. Trata-se de analisar aspectos que contribuem para criar condições mais ou menos favoráveis ao estabelecimento de cadeias referenciais em condições de complexidade linguística. Nas condições que controlámos, há dois tipos de cadeias anafóricas: aquelas em que a correferência se estabelece através de construções relativas e em que o pronome relativo é neutro (QUE) ou possui traços morfológicos de género (O QUAL/A QUAL) ou de animacidade (QUEM) e aquelas em que a correferência é garantida por pronomes pessoais com categorias sintácticas variáveis (demonstrativos vs. pronomes pessoais)

<sup>1</sup> Este trabalho foi realizado no âmbito do projecto PTDC/LIN/67854/2006 *Compreensão na Leitura. Processamento de Palavras, Frases e Textos*, financiado pela Fundação para a Ciência e a Tecnologia.

Nos modelos de processamento de frases em que a sintaxe tem um papel crucial na atribuição da estrutura sobre a qual se faz a interpretação, há a predição de princípios de *parsing* que actuam no sentido da economia de custos cognitivos, com preferência por construções sintacticamente mais simples, ou menos complexas em termos derivacionais e representacionais. No caso particular de cadeias referenciais, os princípios mais produtivos são o *Filler-gap strategy* (Frazier 1987; Frazier e Clifton 1989), para o tratamento de cadeias com WH-, e o *Minimal Chain Principle* (De Vincenzi 1991) para o tratamento de cadeias com pronomes pessoais, nulos ou realizados.

Relativamente ao processamento de cadeias com WH-, *Filler-gap strategy* prediz que sempre que é encontrado um *filler*, automaticamente é predita a existência de um lugar vazio na estrutura (um *gap*); o *filler* fica activo em memória de trabalho e o sistema de processamento procura activamente o *gap* onde esse *filler* deve ser ligado. Quanto mais rapidamente for realizada a ligação *filler-gap*, menos custos haverá para o sistema. Em consequência, várias predições se poderão fazer sobre o que pode funcionar como pista linguística para que esta associação se possa fazer com mais ou menos eficiência.

No caso de cadeias com pronomes pessoais, nomeadamente pronomes nulos (pro) ou foneticamente realizados, De Vincenzi (1991) propôs o *Minimal Chain Principle*, que prediz que o sistema de processamento de frases deve evitar postular nós não necessários na estrutura, mas que não deve adiar fazê-lo se forem necessários.

## 1. O Estudo

Partindo do conceito de complexidade linguística e dos seus efeitos na compreensão, realizámos um estudo exploratório em que, em vez de canonicamente partir de estruturas construídas e testadas em isolamento de quaisquer outras variáveis para além daquelas que se pretendem controlar – os usuais testes de processamento de frases isoladas – optámos por texto real paradigmaticamente considerado de alto nível de complexidade (sentido comum) e, aí, identificámos as construções referenciais que queríamos observar. São objectivos do estudo:

- (i) Identificar, em situação real de leitura, indicadores de custos de processamento na integração de cadeias referenciais com propriedades gramaticais distintas.
- (ii) De entre um conjunto de variáveis linguísticas que criam restrições ao processamento referencial, avaliar a sua validade no processamento do Português Europeu e seleccionar as mais interessantes para posterior estudo em situação canonicamente controlada.

## 2. Metodologia

### *Participantes*

Participaram neste estudo voluntariamente 20 alunos da Faculdade de Letras da Universidade de Lisboa, todos com visão normal ou corrigida. Todos os participantes eram falantes nativos de Português Europeu e têm por hábito ler em Português, livros de literatura e outros, pretendendo-se reunir indivíduos com perfil de “leitor de livros” (veja-se o estudo *A Leitura em Portugal*, Lima *et al.* 2007).

Com o objectivo de avaliar o efeito da habituação a certos padrões de escrita, na amostra seleccionada, registámos dados sobre o perfil de leitor, que nos permitem constituir dois subgrupos: um de leitores e outro de não-leitores do autor do texto utilizado como estímulo, José Saramago. Esta informação permitir-nos-á posteriormente contrastar resultados com estudos informais já realizados e, de forma mais sistemática, estudar o efeito da exposição a certos estilos de escrita marcados por uma retórica particular.

### *Procedimento*

O equipamento de registo utilizado foi o modelo R6-HS da ASL, um sistema remoto monocular com uma resolução de 240Hz que regista os comportamentos oculares com base no reflexo da pupila e da córnea, ou antes, com base na posição e distância entre estes dois pontos de reflexão. Para minimizar os movimentos de cabeça, o que diminui a acuidade do registo, utilizámos um apoio de queixo.

O texto foi dividido em partes, por 10 *slides*, a Verdana a 16 pontos, e apresentado num ecrã de computador de 17 polegadas colocado a 60 cm dos participantes. Para apresentar os estímulos e para controlar o início e o fim da gravação dos comportamentos oculares, que é realizada pelo programa Eye-Trac 6000, recorremos ao programa E-Prime. O E-Prime, comunicando com o Eye-Trac 6000, permite dar início automaticamente à gravação quando o estímulo aparece no ecrã e terminar quando este desaparece.

Antes de dar início à experiência, foram dadas indicações sobre a postura a adoptar pelos participantes, sobre a forma de proceder à calibração da máquina e também ao modo como deveriam fazer avançar o texto no ecrã. Por fim, os participantes foram informados, oralmente pelo observador e posteriormente por escrito, nas instruções que liam no monitor, de que a tarefa consistia em ler para compreender e que no final teriam de responder a algumas questões sobre o texto lido.

Depois do processo de calibração, os participantes liam um *slide* de texto com dados biográficos de José Saramago que servia para verificar se o equipamento estava bem calibrado e para se acostumarem à tarefa.

Posteriormente à realização da experiência, os informantes preencheram uma ficha de dados, caracterizadora do perfil de leitor do participante, e responderam por escrito ao questionário sobre o texto lido.

### *Materiais e estruturas controladas*

Utilizámos como material experimental um excerto do livro *O Conto da Ilha Desconhecida*, de José Saramago. Tendo como objectivo analisar os custos de processamento de cadeias de referência, seleccionámos diferentes estruturas em que uma entidade é introduzida e posteriormente retomada. As expressões anafóricas utilizadas para as retomas são variadas, desde pronomes relativos a pronomes demonstrativos ou nulos, como se verá adiante. Para além de diferentes expressões anafóricas, os contextos em análise têm propriedades sintácticas, morfológicas e lexicais distintas que, por serem inseridos em texto real, não foram controladas ou manipuladas. A descrição linguística mais detalhada de cada uma das estruturas será feita ao longo da análise e discussão dos resultados.

As estruturas seleccionadas, assinaladas a *itálico*, assim como as respectivas representações sintácticas são as apresentadas de (1) a (4); de forma simplificada, iniciamos cada representação apenas com o NP que refere a primeira entidade introduzida que inicia a cadeia ou série de cadeias em análise.

- (1) */o primeiro-secretário chamava o segundo-secretário<sub>i</sub>, leste<sub>i</sub> chamava o terceiro<sub>j</sub>, lque<sub>j</sub> mandava o primeiro-ajudante<sub>k</sub>, lque<sub>k</sub> por sua vez mandava o segundol*  
 [NP<sub>i</sub> [CP(PRON(este)<sub>i</sub> VO<sub>j</sub>) [RC (que<sub>j</sub> VO<sub>k</sub>) [RC(que<sub>k</sub> (parentética) [v]<sub>k</sub> VO)
- (2) */até chegar à mulher da limpeza<sub>i</sub>, /a qual<sub>i</sub>, não tendo ninguém<sub>j</sub> /em quem<sub>j</sub> mandar, / entreabria a porta/... /e perguntava pela frincha/*  
 [NP<sub>i</sub> [RC(a qual<sub>i</sub> [CP V [pro]<sub>i</sub> O<sub>j</sub>) [RC(em quem<sub>j</sub> [PRO]<sub>i</sub> V) [v]<sub>i</sub> V O [CP [conj] (e)] [pro]<sub>i</sub> V PP
- (3) */um parecer fundamentado por escrito ao primeiro-secretário<sub>i</sub>, /o qual<sub>i</sub>, escusado seria dizer, / passava a encomenda ao segundo-secretário<sub>j</sub>, /este<sub>j</sub> ao terceiro, /*  
 [NP<sub>i</sub> [RC(o qual<sub>i</sub> (parentética) [v]<sub>i</sub> V O Obl<sub>j</sub>) [CP(PRON (este<sub>j</sub> [v] Obl
- (4) */até chegar outra vez à mulher da limpeza<sub>i</sub>, /que<sub>i</sub> despachava sim ou não /conforme estivesse de maré./*  
 [NP<sub>i</sub> [RC(que<sub>i</sub> V [CP[[-]<sub>i</sub> V

As quatro condições não apresentam uniformidade quanto às cadeias que integram. Na condição (1), temos uma frase em que o primeiro nexos referencial ocorre por uso do demonstrativo ESTE, que inicia uma frase aposta (ou assindética); as entidades subsequentes são retomadas por relativização e uso do morfema relativo QUE. Na condição (2), há uma cadeia referencial dominante, cuja cabeça é *a mulher da limpeza*, retomada pelo morfema relativo A QUAL, e por pronomes nulos; encaixada nessa cadeia, ocorre uma outra por introdução de nova entidade formada por relativização, em que o constituinte relativo EM QUEM tem a função sintáctica de Oblíquo. A condição (3), semântica e sintacticamente, apresenta um certo paralelismo com a condição (1), embora seja menos complexa em termos de entidades retomadas: a primeira é retomada pelo relativo O QUAL, e a segunda pelo demonstrativo ESTE, numa estrutura sintáctica paralela a (1); note-se que há a inserção de material parentético, *escusado seria dizer*, entre o morfema relativo/sujeito e o verbo, criando uma descontinuidade que poderá ter implicações no processamento da cadeia referencial. Finalmente, na condição (4), retoma-se a mesma entidade da condição (2), *a mulher da limpeza*, que aqui é retomada pelo relativo QUE e por pronome nulo na frase adverbial final.

Há ainda um tipo de categoria sintáctica que tem sido considerada na investigação, estudada no âmbito do processamento de categorias vazias (Fodor 1993; Barss 1993). É o caso dos vestígios deixados por movimento do constituinte relativo que, nas condições que estudamos, se tornam interessantes em situação em que há interposição de material “parentético” entre o morfema relativo e o verbo. Nas condições acima apresentadas, temos três ocorrências a assinalar: em (1): *que por sua vez* [v] *mandava*

o segundo; em (2): a qual, não tendo ninguém em quem mandar [v] entreabria a porta; e em (3): o qual, escusado seria dizer, [v] passava....

### Hipóteses

Pelas propriedades semânticas e sintáticas que exibem, as estruturas insertas nas quatro condições acima referidas poderão apresentar níveis de complexidade diferente em consequência: (i) de representarem uma só cadeia ou várias cadeias referenciais; (ii) do estatuto sintático e morfológico das expressões de retoma anafórica; (iii) dos processos sintáticos de ligação de orações que interagem com a criação da cadeia referencial. Ponderadas as propriedades referidas, e considerando efeitos de localidade sintáctica nas relações dependenciais (*Syntactic prediciton locality theory*, Gibson 1998) e as relações entre os tipos de termos anafóricos e o suposto grau de activação do antecedente que designam (*The accessibility Marking Scale*, Ariel 1996), poderemos prever que:

- (i) Pronomes foneticamente não realizados, ao retomarem a entidade mais proeminente, sujeito e tópico na frase precedente, devem criar condições mais favoráveis do que qualquer outro tipo de expressões anafóricas.
- (ii) Demonstrativos que retomam antecedentes contíguos criam condições óptimas para o processamento do antecedente.
- (iii) Expressões WH- marcadas morfológicamente com traços de género, número ou animacidade deverão favorecer a ligação com o antecedente, mais facilmente do que no caso de WH- não marcado.
- (iv) Expressões WH- com função de sujeito, mantendo a ordem canónica da frase, criam condições mais favoráveis do que quando possuem outras funções sintáticas.

Em termos metodológicos, recorremos ao registo dos movimentos dos olhos, por se esperar que os custos de processamento impostos pela resolução das cadeias de referência tenham correlatos nos padrões de percepção visual, quer ao nível do número e da duração das fixações oculares, quer quanto à direcionalidade dos movimentos sacádicos.

### Tratamento dos dados

Nos excertos apresentados, criaram-se áreas de interesse para cada um dos contextos seleccionados (assinalados nas estruturas de (1) a (4) com barras verticais). Assim, de todo o percurso do olhar realizado durante a leitura do excerto, analisaram-se apenas os comportamentos referentes à leitura das estruturas alvo. Observaram-se como variáveis dependentes o tempo total de leitura calculado em função da soma das fixações, o número de fixações e a sua duração média. Uma vez que os contextos apresentam extensões bastante distintas, optámos por utilizar como medida mais fiável o tempo por carácter.

(1)	o primeiro- secretário chamava o segundo- secretário,	este chamava o terceiro,	que mandava o primeiro- ajudante,	que por sua vez mandava o segundo,
	<b>SN<sub>i</sub></b>	<b>DEM<sub>i</sub></b>	<b>WH<sub>j</sub></b>	<b>WH<sub>k</sub></b>
Num.Carcteres	51	24	31	34
Tempo.Caracter	<b>0.041</b>	<b>0.057</b>	<b>0.046</b>	<b>0.053</b>
	0.013	0.018	0.015	0.021
Dur.MedFix	<b>0.258</b>	<b>0.276</b>	<b>0.264</b>	<b>0.253</b>
	0.054	0.059	0.046	0.054
Num.Fixações	<b>8.3</b>	<b>5.1</b>	<b>5.4</b>	<b>7.0</b>
	2.7	1.7	1.6	1.9

(2)	até chegar à mulher da limpeza,	a qual, não tendo ninguém	em quem mandar,	entreabria a porta	e perguntava pela frincha,
	$SN_i$	$WH_i$	$WH_j$	$pro_i V$	$pro_i$
<i>Num.Carcteres</i>	31	26	15	19	26
<i>Tempo.Character</i>	<b>0.043</b>	<b>0.061</b>	<b>0.058</b>	<b>0.055</b>	<b>0.054</b>
	0.015	0.030	0.029	0.027	0.019
<i>Dur.MedFix</i>	<b>0.275</b>	<b>0.266</b>	<b>0.246</b>	<b>0.355</b>	<b>0.254</b>
	0.075	0.059	0.055	0.139	0.054
<i>Num.Fixações</i>	<b>5.0</b>	<b>6.0</b>	<b>3.5</b>	<b>3.2</b>	<b>5.6</b>
	1.9	2.6	1.6	1.7	1.8

  

(3)	um parecer fundamentado por escrito ao primeiro- secretário,	o qual, escusado seria dizer,	passava a encomenda ao segundo- secretário,	este ao terceiro,
	$SN_i$	$WH_i$	$pro_i V$	$DEM_j$
<i>Num.Carcteres</i>	58	29	41	16
<i>Tempo.Character</i>	<b>0.046</b>	<b>0.050</b>	<b>0.049</b>	<b>0.034</b>
	0.016	0.020	0.018	0.021
<i>Dur.MedFix</i>	<b>0.288</b>	<b>0.254</b>	<b>0.262</b>	<b>0.288</b>
	0.053	0.077	0.047	0.117
<i>Num.Fixações</i>	<b>9.3</b>	<b>6.3</b>	<b>7.8</b>	<b>2.1</b>
	2.4	3.4	2.5	1.1

  

(4)	até chegar outra vez à mulher da limpeza,	que despachava sim ou não	conforme estivesse de maré.
	$SN_i$	$WH_i$	$pro_i$
<i>Num.Carcteres</i>	30	25	26
<i>Tempo.Character</i>	<b>0.034</b>	<b>0.046</b>	<b>0.061</b>
	0.010	0.022	0.020
<i>Dur.MedFix</i>	<b>0.234</b>	<b>0.259</b>	<b>0.290</b>
	0.038	0.065	0.066
<i>Num.Fixações</i>	<b>6.2</b>	<b>4.6</b>	<b>5.8</b>
	1.9	2.2	1.9

**Tabela 1** – Valores de leitura nas condições seleccionadas. (*Num.Caracteres* – número de caracteres de cada contexto de análise; *Tempo.Character* – relação do tempo total de leitura do contexto pelo número de caracteres; *Dur.MedFix* – duração média das fixações; *Num.Fixações* – número de fixações realizadas no contexto. A negrito assinalam-se os valores médios e a cinza o valor do desvio padrão.)

### 3. Análise e Discussão de Resultados

Calculado o tempo global por carácter, considerando a leitura dos vinte sujeitos e dos 16 contextos analisados, chegou-se ao valor de 49ms/carácter, com um desvio padrão de 0.022; a duração média da fixação ficou em 270ms, com o desvio padrão de 0.072 (note-se que, neste último caso, encontramos valores muito próximos aos verificados em Luegi (2006) na leitura de textos de escrita corrente, não literária: 286ms/fixação). Observados os valores mais altos e mais baixos de tempo por carácter, encontrados na Tabela 1, verifica-se que há dois contextos com 61ms – *a qual, não tendo ninguém* (condição 2) e *conforme estivesse de maré* (condição 4), e dois contextos com 34ms – *este ao terceiro*, (condição 3) e *até chegar outra vez à mulher da limpeza*, (condição 4). A análise destes resultados far-se-á considerando as cadeias que integram.

Analisada cada uma das condições, verificamos que há tendência para uma subida de tempo entre o contexto que introduz uma nova entidade e o contexto subsequente em que é retomada por uma qualquer expressão anafórica. Isto confirma-se na variação do tempo de leitura do primeiro para o segundo contexto nas quatro condições. No entanto, no interior de cada condição, verifica-se que há variações. Em (1), onde o NP/Objecto de cada oração é recursivamente modificado por uma frase



relativa, temos uma redução de tempo, no terceiro contexto onde se retoma a terceira entidade, seguida de uma subida no último contexto: (41ms), 57ms, 46ms e 53ms. Em (3), onde há um certo paralelismo por via da estrutura frásica e do léxico, os valores vão no sentido de uma manutenção nas segunda e terceira retomas anafóricas e de uma surpreendente descida no último contexto: (46ms), 50ms, 49ms e 34ms. Já em (2), onde há uma cadeia referencial uniforme com manutenção da mesma entidade apenas interrompida pela introdução de uma segunda entidade menos definida (*ninguém*), os tempos observados decrescem: (43ms), 61ms, 55ms e 54ms. Ao contrário, na condição (4), em que também há uma cadeia referencial centrada numa mesma entidade, os valores vão em *crescendum*: (34ms), 46ms e 61ms.

Em suma, nada se poderá concluir quanto a uma uniformidade na tendência de uma subida ou descida de tempo no processamento da referência que possa ser imputada à ordem de introdução das entidades no discurso e à sua retoma anafórica em cadeias referenciais. Lembrando as condições de leitura, note-se que a primeira e segunda condição ocorrem no segundo *slide* e a terceira e a quarta no terceiro *slide* (foram lidos 10 *slides*), fazendo com que em (3) e (4) se refiram entidades já conhecidas e reintroduzidas por estruturas relativamente próximas.

Tentando ver o que se passa em termos de variações de tempo de leitura, retomamos os tempos mais desviantes relativamente à média encontrada. Os valores mais altos encontrados, como já referido, são os referentes ao segundo contexto da condição (2) e ao último da condição (4). Não devendo imputar-se a subida de tempos à morfologia da expressão anafórica, no primeiro caso o morfema portador de marcas de género QUAL, e, no segundo caso, um pronome nulo para designar uma entidade que se mantém como tópico na frase complexa, deverão ser consideradas as propriedades das construções respectivas. Em *a qual, não tendo ninguém em quem mandar* (valor corrigido médio para os dois contextos de 60ms), temos uma construção sintáctica complexa em consequência da adjunção à frase matriz de uma oração adverbial gerundiva negativa, que se insere depois do pronome relativo/sujeito, e que possui um sujeito pronominal nulo correferente com a cabeça da cadeia referencial em construção que, a realizar-se foneticamente, ocorreria em posição pós-verbal. Em *conforme estivesse de maré*, temos uma frase final adverbial (conformativa) adjunta ao VP da oração relativa precedente, que possui, também, um sujeito pronominal nulo correferente com a cabeça da cadeia referencial em construção; para além disso, em termos lexicais, integra uma expressão lexicalizada *estar de maré* que pode colocar problemas de acesso ao significado. Estes dois casos, curiosamente, constituem as únicas situações em que temos do ponto de vista sintáctico construções adverbiais que podem colocar problemas de integração na estrutura frásica complexa a que pertencem. Embora em contexto de estudo da relação entre complexidade linguística e produção, Gayraud e Martinie (2008) verificam haver maior dificuldade na produção de frases adverbiais vs. modificadores, de entre os quais frases relativas, e mais dificuldade entre os dois do que na integração de frases completivas. Explicam a variação na dificuldade de processamento pela maior ou menor integração desses constituintes na frase complexa. É pista de reflexão que fica para explorar.

Analisadas as estruturas onde há maior decréscimo de tempo, verificamos o que poderá ser interpretado por um efeito de *priming* sintáctico. Comparem-se os valores encontrados para os contextos: *este chamava o terceiro* (57ms) vs. *este o terceiro* (34ms) e *até chegar à mulher da limpeza* (43ms) vs. *até chegar outra vez à mulher da limpeza* (34ms). Em ambos os casos, a expressão com tempo significativamente mais curto foi lida posteriormente a estruturas quase idênticas. Interpretamos esta leitura rápida com consequência de uma habituação a padrões estruturais e discursivos que facilitam o processamento em termos de decodificação e integração. Note-se que o facto de em *este ao terceiro* haver uma construção de elipse de verbo e de objecto, parece não criar problemas de integração, se negligenciarmos a leve subida da duração da fixação relativamente ao tempo médio (288ms vs. 270ms, respectivamente).

Uma das variáveis que nos parecia poder introduzir variação no processamento referencial e ser indicador de complexidade era o tipo da expressão anafórica, considerando o seu potencial designativo do antecedente. Temos cadeias com pronomes relativos, marcados ou não, pronomes demonstrativos e pronomes nulos. Sendo claro que é difícil isolar o peso da informação morfológica do constituinte relativo do contexto frásico que integra, se contrastarmos os contextos iniciados por QUE (46ms, 53ms e 46ms) contra os iniciados por O QUAL/A QUAL/EM QUEM (61ms, 50ms e 58ms), em termos médios, o primeiro apresenta um valor mais baixo (48ms vs. 56ms). Tais resultados contrariariam a hipótese de que a informação morfológica na expressão anafórica facilitaria a ligação com o antecedente. Contudo é de notar que no caso dos contextos iniciados por A QUAL e O QUAL, há a inserção de material parentético que pesa na estrutura sintáctica e que cria distância entre *filler* e

*gap*. Por outro lado, EM QUEM representa a extracção de um oblíquo com repercussões na ordem de palavras (OBL/SU/V) e numa estrutura em que há uma categoria vazia, um PRO, sujeito de frase infinitiva. Por outro lado, deve pensar-se no estatuto de QUE, pronome ou complementador (Mateus *et al.* 2003), que certamente terá efeitos numa tipologia de WH- com valor referencial.

O demonstrativo ESTE ocorre em duas instâncias com valores completamente díspares. Se é inesperado o alto valor de tempo associado à primeira ocorrência (57ms), é explicável o baixo valor na segunda (34ms) pelo efeito de *priming* já acima invocado. Na escala de Ariel (1996:21), o demonstrativo obviamente surge depois de nulos, reflexos, traços de flexão, de clíticos, de pronomes realizados, mas com valor designativo alto. Não são suficientes as condições nem as frequências de ocorrências para extrairmos conclusões.

Finalmente, os dois pronomes nulos que ocorrem em (2) e em (3) apresentam os valores de 54ms e 61ms, respectivamente, enquanto que os dois vestígios por movimento WH-, em (2) e (3), apresentam os valores de 55ms e de 50ms. Descontando o facto de os 61ms em (3) poderem ser consequência da estrutura sintáctica oracional, não encontramos pistas que permitam distinguir entre categorias (v. Fodor (1993) onde se apresentam evidências sobre o processamento dos vários tipos de categorias vazias).

### Conclusões

O estudo apresentado insere-se num projecto alargado que tem por objectivo global a *Compreensão na leitura* e, particularmente, *o processamento de palavras, frases e textos*. Como já referimos, o presente estudo tem vincado carácter exploratório pelas opções que fizemos em termos dos materiais experimentais que quisemos que fossem naturais e que induzissem uma leitura natural. Esta opção releva de duas preocupações. A primeira, e mais importante, é que pretendemos iniciar uma investigação que conduza ao reconhecimento de indicadores de complexidade linguística com repercussões no processamento, na compreensão e na produção verbal. Se, para tal, possuímos já alguns dados sobre a produção escrita em que trabalhamos sobre dados da produção escrita natural, não controladamente induzida (Costa e Gonçalves, em preparação), para a compreensão de textos não temos ainda dados recolhidos, excepto os que possuímos sobre o processamento de frases isoladas (Costa 2005; Maia *et al.* 2007); sobre o processamento de palavras, ver Faria e Luegi 2008. Exactamente porque para o processamento de frases temos trabalhado em paradigmas experimentais altamente controlados que recorrem a baterias de frases descontextualizadas, onde se manipulam as variáveis sintácticas a controlar, decorre a segunda preocupação: a de experimentar um paradigma de investigação ecologicamente válido, o que é hoje uma preocupação na “era moderna” da Psicolinguística. Na investigação psicolinguística cresce, hoje, o interesse em trabalhar com materiais autênticos que, pela sua complexidade, impõem condições de processamento mais informativas dos processos cognitivos implicados (v. Garnham, Garrod e Sanford 2006).

Apesar de não podermos, porque o material não é directamente contrastável, quer pela sua extensão, variável entre condições, quer pelos diversos factores linguísticos (não controlados) que podem justificar as variações encontradas entre as condições, extrair dos resultados obtidos grandes conclusões, podemos e devemos identificar algumas pistas que podem servir de base a estudos a realizar futuramente. Destacam-se: (i) a diferença de custos de processamento entre a introdução de uma entidade e a sua retoma; (ii) custos de processamento de diferentes expressões referenciais, com características sintácticas próximas; (iii) efeitos de *priming*, ou seja, redução de custos de processamento de uma determinada estrutura quando esta é precedida de outra estruturalmente semelhante; (iv) custos de processamento de estruturas em que se insere material, frases parentéticas, por exemplo, entre o sujeito e o verbo.

Relativamente ao primeiro ponto, os dados indicam que a retoma de uma entidade tem custos mais elevados do que a sua introdução. Seria interessante testar estes resultados de uma forma mais controlada, uma vez que a maioria dos estudos realizados restringem as suas análises ao contraste entre os tipos de retomas, por exemplo, NP pleno *vs.* pronome (veja-se Almor 2000).

O ponto (ii) merecia ser alvo de uma análise mais detalhada, uma vez que, pelas razões atrás apresentadas, pouco se pode concluir dos resultados obtidos. Contudo, seria interessante tentar perceber, com material mais controlado, se existem ou não diferenças significativas entre o processamento de pronomes neutros e pronomes marcados morfológicamente (cf. trabalho em curso sobre processamento cognitivo de pronomes em Português, de P. Luegi). Lembremos que o Português, por poder ter marcação de género, número e animacidade nos morfemas WH-, se apresenta como uma língua vantajosa para esta verificação.



O estudo do efeito de *priming* em texto deveria também ser alvo de uma análise mais aprofundada. Os trabalhos realizados têm-se focado no efeito de *priming* entre frases apresentadas isoladamente e, para além disso, este efeito tem sido sobretudo estudado ao nível da produção (veja-se, por exemplo, Ledoux, Traxler e Swaab 2007). O seu estudo em texto parece por isso fundamental. Por fim, e partindo dos indícios levantados com o trabalho que agora apresentamos, parece necessário investigar com mais rigor os custos associados ao processamento de vestígios, sobretudo em casos em que entre *filler* e *gap* ocorre a inserção de material. Nesse caso, será fundamental controlar a estrutura sintáctica e propriedades lexicais do tipo de material que ocorre entre o sujeito – expressão referencial – e o verbo.

Como nota final, e apesar dos problemas que fomos levantando ao longo da apresentação e discussão dos resultados, queremos reforçar a necessidade de realizar estudos deste tipo, em que se recorre a texto real. Para além dos resultados que daí advêm, muitas questões podem justamente surgir quando se analisa com maior detalhe os comportamentos dos informantes perante este tipo de material. Muitas questões com que nos deparámos ao longo deste estudo não teriam surgido se tivéssemos recorrido a material construído para a análise de uma variável em particular e por isso parece-nos fundamental prosseguir com este género de trabalhos.

### Referências

- Almor, A. (2000). Constraints and mechanisms in theories of anaphor processing. In M. W. Crocker, M. Pickering, & C. Clifton, *Architectures and Mechanisms for Language Processing* (pp. 341-354). Cambridge, UK: Cambridge University Press.
- Ariel, M. (1996). Referring expressions and the +/- coreference distinction. In J. Gundel, & T. Fretheim, *Referent and Referent accessibility* (pp. 13-35). John Benjamins.
- Costa, M. A. (2005). *Processamento de frases em Português Europeu. Aspectos cognitivos e linguísticos implicados na compreensão da língua escrita*. Lisboa: Fundação Calouste Gulbenkian.
- Costa, M. A. (2006). Processamento de frases declarativas transitivas com SU pós-verbal - padrões de ordem emergentes. In (falta o nome dos organizadores) *Revista Letras de Hoje*, 143, Porto Alegre, Rio Grande do Sul: PUC do Rio Grande do Sul, 235-261.
- Costa, A., Faria, I. F., Kail, M. (2004). Semantic and Syntactic Cues' Interaction on Pronoun Resolution in European Portuguese. In Branco, McEnery & Mitkov (Eds.) *DAARC 2004, 5<sup>th</sup> Discourse Anaphora Resolution Colloquium*. Proceedings. Lisboa: Ed. Colibri, 45-50.
- Costa, M. A., Faria, I. H., Matos, G. (1997). Ambiguidade referencial na identificação do sujeito em estruturas coordenadas. *Actas do XI Encontro da APL*. 1997. Lisboa.
- De Vincenzi, M. (1991). *Syntactic Parsing Strategies in Italian: The Minimal Chain Principle*. Dordrecht: Kluwer Academic Publishers.
- Faria, I. H., Luegi, P. (2008a). Priming effect on word reading and recall. In Antonis Botinis (Ed.), *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics*. ISCA and University of Athens: Greece, pp. 89-92.
- Frazier, L., Charles, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4, 93-126.
- Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory* 5, 519-559.
- Gayraud, F., Martinie, B. (2008). Does Structural Complexity Necessarily Imply Processing Difficulty? *Journal of Psycholinguist Research*, 37:21-31
- Gibson, Edward (1998). Linguistic complexity: locality of syntactic dependencies, *Cognition* 68, 1-76.
- Kail, M. (1989). Cue validity, cue cost, and processing types in sentence comprehension in French and Spanish. In MacWhinney, B. e Bates, E. (Eds.). *The cross-linguistic study of sentence processing*. NY: University Press.
- Ledoux, K., Traxler, M., Swaab, T. (2007) Syntactic priming in comprehension: evidence from event-related potentials. *Psychol Sci* 18(2):135-43.
- Leitão, M. (2005). *O Processamento do objeto direto anafórico no Português Brasileiro*. Tese de Doutoramento apresentada à Universidade Federal do Rio de Janeiro . Rio de Janeiro.
- Lima, M. L. S. (Coord.) (2007). *A Leitura em Portugal*. Lisboa: GEPE-ME.
- Luegi, P. (2006). *O Registo do Movimento dos Olhos durante a Leitura de Textos*. Tese de mestrado apresentada à Faculdade de Letras da Universidade de Lisboa.
- Maia, M., Fernández, E., Costa, A., Lourenço-Gomes, M. C. (2007). Early and late preferences in relative clause attachment in Portuguese and Spanish. *Journal of Portuguese Linguistics*.

- Mateus, M. H., M., Brito, A. M., Duarte, I., Faria, I. H., Frota, S., Matos, G., Oliveira, F., Vigário, M., Villalva, A. (2003). *Gramática da Língua Portuguesa*. Lisboa: Caminho. [5.<sup>a</sup> ed. revista e aumentada]
- Vasconcelos, M. (1996). Compreensão e produção de frases relativas em Português Europeu. Faria *et al.* (Org). *Introdução à Linguística Geral e Portuguesa*. Lisboa: Caminho.
- Warren, T., Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, 85, 79-112.